CONDITIONALLY ACCEPTABLE FREQUENTIST SOLUTIONS[1]

George Casella

Biometrics Unit
Cornell University
Ithaca, New York 14853, U.S.A.

I. INTRODUCTION

Statistical procedures constructed according to traditional frequentist criteria alone can suffer from poor conditional behavior. A conditional frequentist criterion, based on a theory of relevant betting strategies, can provide an objective framework for assessing conditional performance. Previous authors have advocated conditional confidence principles based on this theory, and it is argued here that even stronger conditional criteria are needed. Some examples are given to substantiate this claim, and procedures which satisfy the stronger conditional requirement, and hence are conditionally acceptable, are given.

A *frequentist solution* is a solution to a statistical problem that has associated with it some guaranteed long-run property. In particular, for the problem of interval estimation a frequentist solution is a confidence interval that has a minimum coverage guarantee. If $X|\theta \sim f(X|\theta)$, a procedure $C(X)$ is a frequentist solution to the interval estimation problem if for some α, $0<\alpha<1$,

$$P_\theta[\theta \in C(X)] \geq 1-\alpha \quad \forall\theta ,$$

where

$$P_\theta[\theta \in C(X)] = \int_{\mathcal{X}} I[\theta \in C(x)] f(x|\theta)dx .$$

Why are frequentist solutions desirable? The obvious reason is that the long-run guarantee is comforting — one gets some assurance of repeatability of results. Another nice feature is that the frequentistsolution

[1]Research partially supported by NSF Grant MCS-8501973. This paper is BU-908-M in the Biometrics Unit Series, Cornell University.

gives a pre-experimental guarantee: one can set-up an experiment to yield a required precision, and know the effort required to get such precision.

Problems arise with frequentist solutions post-experimentally, however, in that a perceived post-experimental precision can be quite different from the frequentist pre-experimental precision. Here are two examples:

Example 1. The Seidenfeld-Mayo Controversy

Seidenfeld and Mayo, two Philosophers of Inference, have argued about the following set-up:

Let $X|\theta \sim$ Uniform $(0,\theta)$. The best (Neyman-Shortest) $1-\alpha$ confidence interval, based on one observation $X=x$, is given by

$$x \le \theta \le x/\alpha.$$

Yet, if it is known that $\theta \le 15$, and $X=10$ is observed, then with $\alpha=.05$ the realized 95% interval is

$$10 \le \theta \le 200,$$

which, of course, has a 100% probability of containing the true θ. Note that truncating the interval at $\theta=15$ does not remedy this discrepancy.

Example 2. (Berger and Wolpert, 1984) Let X_1, X_2 be independent with

$$P_\theta(X_i = \theta-1) = P_\theta(X_i = \theta+1) = \tfrac{1}{2}, \quad i = 1,2.$$

The confidence set

$$C(x_1,x_2) = \begin{cases} \text{the point } \tfrac{1}{2}(x_1+x_2) & \text{if } x_1 \neq x_2 \\ \text{the point } x_1-1 & \text{if } x_1 = x_2 \end{cases}$$

satisfies

$$P_\theta[\theta \in C(X_1,X_2)] = .75 \quad \forall\theta,$$

but

$$P_\theta[\theta \in C(X_1,X_2) \mid X_1 = X_2] = .5 \quad \forall\theta$$

and

$$P_\theta[\theta \in C(X_1,X_2) \mid X_1 \neq X_2] = 1 \quad \forall\theta .$$

These examples illustrate that "good" frequentist solutions can have peculiar post-data interpretations. By only concentrating on long-run properties, aberrant per-trial behavior may be masked. One solution to the elimination of aberrant per-trial behavior is the augmentation of frequency theory with a theory of conditional inference. This idea is not new, and different types of conditional inference have been exposed by many authors (e.g., Fisher, Buehler, Kiefer, Robinson, Bondar, Fraser). It will be seen that the betting-based theory of Buehler (1959), formalized by Robinson (1979a,b) provides a reasonable augmentation of frequentist theory.

II. CONDITIONAL INFERENCE

Robinson's formalization of Buehler's ideas is both rigorous and general. While some rigor will be maintained, the situation considered here will be somewhat specialized. Most ideas, however, will carry over to a more general setting.

We suppose that a series of trials takes place where a random variable $X \sim f(x|\theta)$ is observed, where $f(x|\theta)$ is a known density function. Given $X=x$, the frequentist constructs a confidence region, $C(x)$, for θ, with confidence coefficient $1-\alpha$. (Robinson allows $1-\alpha$ to depend on x, which could also be done here, but right now that is tangential to the current concern. The quantity $1-\alpha$ is the pre-experimental confidence, and the present concern is with its post-experimental validity. That is, can the *frequentist* confidence coefficient have a meaningful conditional interpretation?) Seeing x and $C(x)$, a bettor is allowed to place bets, at odds $1-\alpha:\alpha$, as to whether or not θ is covered. The bettor places his bets according to a betting function $s(X)$, which will be taken here to be a signed indicator function, with + denoting a bet for coverage and - denoting a bet against coverage. (Robinson allows $s(X)$ to be any bounded function.) For example, the betting strategy $s(X) = -I(X\in A)$ says to bet against coverage if $X\in A$, the bettor risking α to win $1-\alpha$.

Definition: A betting strategy $s(X)$ is *relevant* for the confidence procedure $\langle C(X), 1-\alpha\rangle$, if for some $\varepsilon>0$

$$E_\theta\{I[\theta\in C(X)] - (1-\alpha)\}s(X) \geq \varepsilon E_\theta|s(X)|, \qquad \forall\theta .$$

In terms of betting strategies that are signed indicator functions, a relevant betting strategy satisfies either

$$P_\theta[\theta \in C(X) | X \in A] \geq 1 - \alpha + \varepsilon, \ \forall\theta \qquad \text{(positive bias)}$$

or

$$P_\theta[\theta \in C(X) | X \in A] \leq 1 - \alpha - \varepsilon, \ \forall\theta \qquad \text{(negative bias)} \ ,$$

showing that a relevant *subset* identifies a portion of the sample space where the conditional coverage probability can be bounded uniformly (in θ) away from $1-\alpha$.

Before discussing confidence principles based on relevant betting, let us first examine why this set-up is a reasonable way to augment frequency theory. First, it is totally objective, requiring no prior input, so it falls within the overall bounds of frequentist inference. Secondly, it provides a means to assess per-trial performance of frequentist procedures, and also allows us to see how bad (or good!) $1-\alpha$ is as a post-data measure of precision. Thirdly, the work of Robinson, Bondar (1977), and Pierce (1973) point the way to constructing procedures that are free of relevant betting strategies. (Such procedures must in some sense be Bayes procedures, possibly against improper priors; formal conditions are given both by Robinson and Pierce. Proper Bayes procedures, however, in general cannot yield a frequentist guarantee, so only improper priors can be considered by a conditional frequentist.)

Conditional confidence principles based on the existence of relevant betting strategies have been put forth by Robinson (1976) and Bondar (1977), among others. In particular, we have:

Robinson's Principle: use no procedure for which a negatively-biased relevant strategy exists.

The logic behind this principle is reasonably straightforward. The statistician generally does not worry to much about being conservative, so positively biased betting is really not that much of a concern. The example of Seidenfeld essentially points out a positively biased set,

Example 1 (Continued). If $X \geq 3/4$, then $X/.05 \geq 15$. For the .95 confidence interval $(X, X/.05)$, if $\theta \leq 15$ then

$$P_\theta[\theta \in (X, X/.05) | X \geq \tfrac{3}{4}] = 1.$$

showing that the set $\{X : X \geq \frac{3}{4}\}$ is positively biased.

Robinson's principle can be criticized on logical grounds, and as a stand-alone foundational principle it fails. This is because one can always take an ultra-conservative procedure to satisfy Robinson's principle: The procedure $\langle C(X), 1-\alpha\rangle = \langle(-\infty,\infty), 0\rangle$ allows no negatively biased relevant betting. Thus, Robinson's principle should only be applied to procedures that are "good" unconditional frequentist procedures.

Given these qualifications, the behavior of C(X) in Example 1 is not particularly troubling in practice. On the other hand, Example 2 gives us a negatively biased bet:

Example 2 (continued). Since

$$P_\theta[\theta \in C(X_1, X_2) \mid X_1 = X_2] = .50 \qquad \forall\theta,$$

assigning a confidence coefficient of .75 to $C(X_1, X_2)$ will allow the bettor the negatively biased relevant strategy $s(X_1, X_2) = -I(X_1 = X_2)$.

This situation is quite distressing, showing that if the data fall in the conditioning set, the coverage probability must be lower than the nominal value. The procedure is not conditionally acceptable.

III. CONDITIONAL ACCEPTABILITY

The elimination of negatively biased relevant betting insures that a procedure will not suffer from any serious conditional flaws. However, this requirement is not quite strong enough to eliminate all aberrant behavior. Consider the following:

Example 3. Stein's Two-Stage Procedure (1947)

Stein showed how to construct a fixed-width confidence interval for a normal mean when the variance is unknown. The procedure is two-stage, and can be described as follows: Fix α, ℓ, and n_0 then

1. Take a sample of size n_0, $x_1, \cdots, x_{n_0}$, from a population with density $n(\mu, \sigma^2)$, and calculate s^2, the sample variance, in the usual way. The final sample size is given by

$$n = \max\left\{ \left[\frac{s^2 t^2}{\ell^2}\right] + 1,\ n_0 + 1 \right\},$$

where $[q]$ denotes the greatest integer less than q, and $t = t_{n_0-1,\alpha/2}$ is the upper $\alpha/2$ cutoff from Student's t with n_0-1 degrees of freedom.

2. Take additional observations $x_{n_0+1},\cdots,x_n$ and choose positive constants $a_1,\cdots,a_n$ such that

$$\Sigma a_i = 1, \qquad a_1 = a_2 = \cdots = a_{n_0}$$

$$s^2 \sum_1^n a_i^2 = (\ell/t)^2 \ .$$

(Such constants always exist.) The interval

$$\sum_1^n a_i x_i \pm \ell$$

then has coverage probability $1-\alpha$ uniformly in μ and σ^2.

There are a few things to note about this procedure:

1. One will always go on to the second stage, but if s^2 is small, i.e., $s^2 \le n_0(\ell/t)^2$, only one additional observation is taken. This can be considered stopping at the first stage.
2. Stein also described a modification of this procedure, using a slightly different stopping rule and eliminating the a_i's. The final interval was $\bar{x} \pm \ell$, where $\bar{x} = \frac{1}{n}\sum_1^n x_i$. The coverage probability of this interval is at least $1-\alpha$ for all μ and σ^2. Although this interval is more attractive from a practical viewpoint it is harder to work with theoretically.

The logic behind Stein's interval is straightforward: if the variance estimate is small, the true variance should be small and one doesn't need additional observations to attain a desired accuracy. Also, a theorem of Robinson (1976) can be applied to show that there are no relevant betting strategies. One should be concerned, however, with the conditional performance of Stein's procedure given that you stop at the first stage (or take just one more observation). On this prominent conditioning set, the procedure's performance is not very good.

Theorem: Let C_S denote the Stein Two-Stage Interval, and $A = \{n = n_0+1\}$, that is, A is the event that the procedure takes one additional observation, essentially stopping at the first stage. Then the conditional probability of coverage, $P_{\sigma^2}(\mu \in C_s | A)$, is independent of μ, and is a monotone decreasing function of σ^2, with $P_{\sigma^2=0}(\mu \in C_s | A) = 1-\alpha$. Thus $P_{\sigma^2}(\mu \in C_s | A) \le 1-\alpha$ for all (μ, σ^2).

Proof: Stein showed that the quantity

$$u = (t/\ell) \sum_{i=1}^{n} a_i(x_i - \mu)$$

has Student's t-distribution with n_0-1 degrees of freedom, and can be written as y/s, where $y \sim n(0,1)$ independent of s. The proof of this follows quickly from his clever definition of the constants $a_1, \cdots, a_n$. Using Stein's machinery and the definition of A we can write

$$\begin{aligned} P_{\sigma^2}(\mu \in C \ | A) &= P_{\sigma^2}(|\Sigma\ a_i(x_i - \mu\ )| \le \ell\ | n = n_0+1) \\ &= P_{\sigma^2}\left(|u| \le t \middle| s^2 \le n_0(\ell/t)^2 \right) , \end{aligned}$$

since the event $\{n=n_0+1\}$ is equivalent to the event $\{s^2 \le n_0(\ell/t)^2\}$. Let χ^2_ν denote a chi-squared random variable with ν degrees of freedom. Since we can write $u^2 = (n_0-1)\chi^2_1/\chi^2_{n_0-1}$ with $s^2 = \sigma^2\chi^2_{n_0-1}/(n_0-1)$, we have

$$P_{\sigma^2}(\mu \in C_s | A) = P(\chi^2_1 \le a\chi^2_{n_0-1} | \chi^2_{n_0-1} \le b/\sigma^2) \tag{3.1}$$

where $a = t^2/(n_0-1)$ and $b = n_0(n_0-1)(\ell/t)^2$.

It is straightforward to establish that

$$\lim_{\sigma^2 \to 0} P_{\sigma^2}(\mu \in C_s | A) = P(\chi^2_1 \le a\chi^2_{n_0-1}) = 1-\alpha , \tag{3.2}$$

where the last equality follows from the definition of a. Also, using L'Hopital's rule, we have

$$\lim_{\sigma^2 \to \infty} P_{\sigma^2}(\mu \in C_s | A) = 0 . \tag{3.3}$$

To show that the conditional coverage probability is decreasing in σ^2, we will show that the second derivative is positive at the zeros of the first derivative. This implies that any interior extrema must be a minimum. However, from (3.3), if there is an interior minimum there must also be an interior maximum, which cannot happen. Therefore, there are no interior extrema and the function is monotone decreasing.

We can rewrite the conditional coverage probability as

$$P_{\sigma^2}(\mu \in C_s | A) = T(\sigma^2)/B(\sigma^2) ,$$

where

$$T(\sigma^2) = \int_0^{ab/\sigma^2} P(y/a < \chi^2_{n_0-1} < b/\sigma^2) f_1(y)dy$$

and

$$B(\sigma^2) = P(\chi^2_{n_0-1} \le b/\sigma^2)$$

and $f_\nu(\cdot)$ is the pdf of a χ^2_ν random variable. At the zeros of the first derivative we must have

$$\frac{T(\sigma^2)}{B(\sigma^2)} = \frac{T'(\sigma^2)}{B'(\sigma^2)} ,$$

and straightforward differentiation will show that

$$\frac{T'(\sigma^2)}{B'(\sigma^2)} = P(\chi^2_1 < ab/\sigma^2) .$$

At the zeros of the first derivative, the sign of the second derivative is given by

$$\begin{aligned} &\operatorname{sgn}[B(\sigma^2)T''(\sigma^2) - T(\sigma^2)B''(\sigma^2)] \\ &\quad = \operatorname{sgn}\left[B(\sigma^2)\left(\frac{B''(\sigma^2)}{B(\sigma^2)}T(\sigma^2) + B'(\sigma^2)\frac{d}{d\sigma^2}P(\chi^2_1<ab/\sigma^2)\right) - T(\sigma^2)B''(\sigma^2)\right] \\ &\quad = \operatorname{sgn}\left[B'(\sigma^2)\frac{d}{d\sigma^2}P(\chi^2_1 < ab/\sigma^2)\right] , \end{aligned}$$

where the first equality follows by some straightforward algebra. Since $B'(\sigma^2) < 0$ and $P(\chi^2_1 < ab/\sigma^2)$ is decreasing in σ^2, it follows that the second derivative is always positive at the zeros of the first derivative, and the proof is complete. □

A graph of this probability is given in Figure 1 for $n_0-1 = 5,10,15$. It can be seen that the probability drops off quite rapidly, going from the nominal level of .9 down to .6 for $\sigma=5$.

In the terminology of Robinson (1979a), the set A is negatively biased *semirelevant* betting procedure, which is defined exactly as the relevant procedure except that ε is replaced by zero.

Definition: A betting strategy $s(X)$ is *semirelevant* for the confidence procedure $\langle C(X),1-\alpha\rangle$ if

$$E_\theta\{I[\theta\in C(X)] - (1-\alpha)\}s(X) \ge 0, \qquad \forall\theta .$$

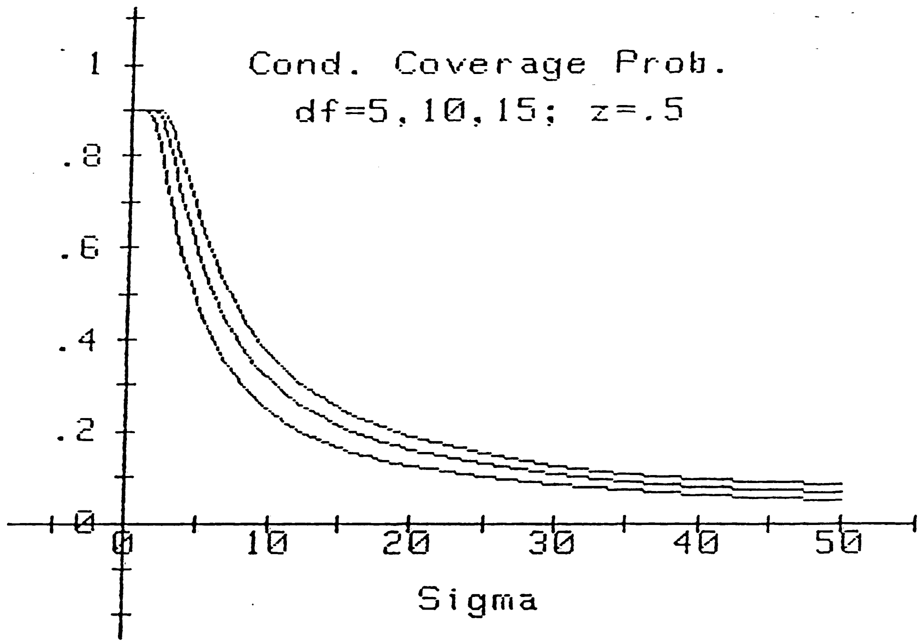

Figure 1: For Stein's two-stage procedure with $1-\alpha = .9$ the probability of coverage, given that you stop at the first stage, is shown. The initial sample sizes are $n_0 = 6,11,16$, and the probabilities are increasing in n_0.

To require procedures to be free of all negatively biased semi-relevant (NBSR) betting is, in many cases, too strong a requirement, and it would eliminate one of the truly great procedures.

Example 4 (Brown, 1967). Let $x_1,\cdots,x_n$ be iid $n(\mu,\sigma^2)$, and let $C(\bar{x},s^2)$ be the usual one-sample $1-\alpha$ Student's t interval. Then there exists constant k such that

$$P_{(\mu,\sigma^2)}[\mu\in C(\bar{x},s^2) \mid |\bar{x}/s| > k] \le 1-\alpha \qquad \forall\mu,\sigma^2$$

Although we cannot eliminate all NBSR betting, we can be selective. More precisely, in any experiment, we can identify sets on which conditional inferences are likely to be made. In the Stein example, or in any multi-stage example, we are faced with inference conditional on stopping early. In Example 4, the set $\{|\bar{x}|/s > k\}$ is the event of rejecting the null hypothesis $\mu=0$, so the above procedure can be interpreted as setting up a confidence interval after a rejection.

The kind of confidence principle that these arguments are leading up to is the following: A procedure is conditionally acceptable if 1) it does not allow negatively biased betting and 2) If A is a set on which it is known that a conditional inference may be drawn (i.e. a stopping region), then A cannot be a negatively biased *semirelevant* subset.

Such a principle eliminates both Stein's two-stage procedure and the t-interval-after-rejection, but does not eliminate the usual t-interval.

IV. CONSTRUCTING CONDITIONALLY ACCEPTABLE SOLUTIONS

Requiring a procedure to be free of negatively biased semirelevant betting on a selected collection of conditioning sets is, in practice, a very cumbersome requirement. Although for some techniques (such as multi-stage procedures) there are obvious conditioning sets (the stopping sets), in general an experimenter would find it difficult to enumerate all sets upon which there is concern over conditional behavior. It is sometimes possible, however, to construct confidence regions that are free of all NBSR betting.

> *Example 5.* Maatta and Casella (1985) are concerned with interval estimation of the variance of a normal population with unknown mean. Two of the intervals examined, C_{ML}, the minimum length interval, and C_{SU}, the shortest-unbiased (or Neyman-shortest) interval, are shown to be free of *all* NBSR betting.

How does one construct a confidence interval that is free of NBSR betting? From the work of Robinson and Pierce we know that Bayes procedures are conditionally acceptable, so what we need do is verify that our candidate procedure is a Bayes credible set with credible probability at least $1-\alpha$.

In particular, suppose $X \sim f(X|\theta)$ and $\langle C(X), 1-\alpha\rangle$ is a $1-\alpha$ confidence procedure. Suppose further that there is a (possibly improper) prior $\pi(\theta)$ for which

$$P_X[\theta\in C(X)] = \int_{\theta\in C(X)} \pi(\theta|X)d\theta \geq 1-\alpha \qquad \forall X , \tag{4.1}$$

where

$$\pi(\theta|X) = \frac{f(X|\theta)\pi(\theta)}{\int f(X|\theta)\pi(\theta)d\theta} .$$

If A is a NBSR subset for $\langle C(X), 1-\alpha\rangle$, i.e., A satisfies

$$P_\theta[\theta \in C(X) | X \in A] - (1-\alpha) \le 0 \qquad \forall\theta \tag{4.2}$$

or equivalently,

$$P_\theta[\theta \in C(X), X \in A] - (1-\alpha)P_\theta(X \in A) \le 0 \qquad \forall\theta ,$$

then multiplying by $\pi(\theta)$ and integrating shows that

$$\int_\Theta \{P_\theta[\theta \in C(X), X \in A] - (1-\alpha)P_\theta(X \in A)\}\pi(\theta)d\theta < 0 . \tag{4.3}$$

Note that this last inequality will be strict as long as (4.2) is strict for some θ getting positive mass from $\pi(\theta)$. If the orders of integration can be interchanged in (4.3), we then have

$$\int_{\mathcal{X}} \{P_X[\theta \in C(X)] - (1-\alpha)\}I(X \in A)m_\pi(X)dX < 0$$

where $m_\pi(X)$ is the marginal distribution of X. Comparing (4.1) and (4.4) gives a contradiction, showing that the set A cannot exist. Therefore, the problem in verifying the absence of NBSR betting reduces to that of finding the prior $\pi(\theta)$ and justifying the interchange of integrals. (Note that since $\pi(\theta)$ will most probably be improper, the interchange of integrals is not automatic.) There is no general prescription for accomplishing this, but it can be carried out in many circumstances.

Example 5 (Continued). Using improper priors of the form

$$\pi(\mu,\sigma) = \frac{1}{\sigma^k} d\mu d\sigma ,$$

it can be shown that C_{ML} and C_{SU} maintain $1-\alpha$ posterior probability.

Example 6. In the problem of estimation of a p-variate normal mean, $p \ge 3$, if $X \sim N_p(\theta, I)$ then the confidence set

$$C^\delta = \{\theta : |\theta - \delta^+(X)| \le c\}$$

has uniformly higher coverage probability than the usual confidence set

$$C^o = \{\theta : |\theta - X| \le c\} ,$$

where $\delta^{+}(X)$ is a positive-part Stein estimator and c satisfies $P(\chi_p^2 \le c^2) = 1-\alpha$ (Hwang and Casella, 1982). C^{δ} is also a posterior Bayes $1-\alpha$ credible set (Casella, 1985). The prior used is an improper prior with the hierarchical form

$$\pi(\theta|\lambda) \sim N_p\left(0, \frac{1-\lambda}{\lambda} I\right)$$

$$\lambda \sim \lambda^{-2}, \qquad 0 < \lambda < 1 .$$

V. DISCUSSION

Problems in conditional inference arise when the pre-experimental and post-experimental precision differ markedly. In statistical practice, we should be most concerned when the pre-experimental confidence is greater than the post-experimental confidence. Robinson's principle, elimination of negatively-biased relevant betting, assures us that there are no gross deviations in the pre- vs. post-experimental confidence. The above augmentation of this principle eliminates more subtle deviations on sets of interest. It should be noted that the above augmentation of Robinson's principle somewhat reflects a concern of Bondar (1977): some conditioning sets may have low probability for most of the parameter space, so we may be worrying about events that are unlikely to occur. However, Bondar formulates a different confidence principle which turns out to be weaker than Robinson's. The principle given here, while addressing this question in asking for sets of interest, leads to a stronger principle than Robinson's.

The construction of conditionally acceptable confidence procedures is not easy, although it has been accomplished in a few settings. The elimination of NBSR betting on a selected collection of sets is a difficult and time consuming task. The blanket elimination of all NBSR betting, as outlined in Section IV, results in procedures that satisfy stronger conditional criteria than required here, but it seems that this blanket elimination is the only practical means of assuring conditional acceptability in practice.

A final question to be addressed is what to do when the pre- and post-experimental measures differ in such a way as to violate the above confidence principle. One can, of course, abandon the procedure, but that is a destructive rather than constructive answer. There doesn't seem to be a constructive answer to cover all situations, but the following illustrates some methods of attack using the Stein two-stage procedure.

Recall that any procedure under consideration does not allow negatively biased relevant betting. If there are sets of interest for which negatively biased semirelevant bets exist, this can be taken as a signal that the pre-experimental confidence is too high. Three possible strategies come to mind:

1. Lower the pre-experimental confidence. (Note that this modification preserves the non-existence of negatively biased relevant bets.) With the pre-experimental confidence lowered, any semirelevant sets can be eliminated. One may choose a pre-experimental confidence based on Figure 1; specifying a maximum sigma will specify a minimum confidence.
2. Quote different pre-experimental confidence for each of the identified negatively biased sets. In the notation of Example 3 and the Theorem,

$$P_{\sigma^2}(\mu \in C_s | A) \le 1 - \alpha \qquad \forall \sigma^2 ,$$

$$P_{\sigma^2}(\mu \in C_s | A^c) \ge 1 - \alpha \qquad \forall \sigma^2 ,$$

   so the post-experimental confidence is greater than $1-\alpha$ on A^c. Quoting a different pre-experimental confidence for A and A^c is a way to eliminate negatively biased semirelevant bets, and, in this example, preserves the non-existence of negatively biased relevant bets.
3. Modify the procedure. For a fixed α and n_0, use $t_{\nu,\alpha/2}$ instead of $t_{n_0-1,\alpha/2}$, where $\nu < n_0-1$. This modification eliminates the negatively biased semirelevant set A, at the expense of increasing the length of the interval.

VI. CONCLUSIONS

Traditional Neyman-Pearson theory contains no mechanism for evaluating conditional performance of its procedures. The theory of relevant betting provides an objective framework for doing this, and is a natural way of augmenting Neyman-Pearson theory to insure that resulting procedures are good both unconditionally and conditionally.

Robinson's principle of conditional inference, to allow no negatively biased relevant betting, does not appear to be strong enough to eliminate all troublesome conditional behavior. The principle outlined in this article is slightly stronger, and leads to procedures with more acceptable conditional performance.

REFERENCES

[1] Berger, J. O., and Wolpert, R. L. (1984). The likelihood principle. *IMS Monograph Series.*

[2] Bondar, J. V. (1977). A conditional confidence principle. *Ann. Statist. 5*, 881-891.

[3] Brown, L. D. (1967). The conditional level of Students' t test. *Ann. Math. Statist. 38*, 1068-1071.

[4] Buehler, R. J. (1959). Some validity criteria for statistical inferences. *Ann. Math. Statist. 30*, 845-863.

[5] Casella, G. (1985). Constructing conditionally acceptable confidence sets. Technical Report No. BU-889-M, Cornell University.

[6] Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*, Third Edition. Hafner, New York.

[7] Fraser, D. A. S. (1977). Confidence, posterior probability, and the Buehler example. *Ann. Statist. 5*, 892-898.

[8] Hwang, J. T. and Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist. 10*, 868-881.

[9] Kiefer, J. (1977). Conditional confidence statements and confidence estimators (with discussion). *J. Amer. Statist. Assoc. 72*, 789-827.

[10] Maatta, J. and Casella, G. (1985). Conditional properties of interval estimators of the normal variance. Technical Report BU-886-M, Cornell University.

[11] Mayo, D. G. (1981). In defense of the Neyman-Pearson theory of confidence intervals. *Philosophy of Science 48*, 269-280.

[12] Pierce, D. A. (1973). On some difficulties with a frequency theory of inference. *Ann. Statist. 1*, 241-250.

[13] Robinson, G. K. (1976). Properties of Student's t and of the Behrens-Fisher solution to the two means problem. *Ann. Statist. 4*, 963-971.

[14] Robinson, G. K. (1979a). Conditional properties of statistical procedures. *Ann. Statist. 7*, 742-755.

[15] Robinson, G. K. (1979b). Conditional properties of statistical procedures for location and scale parameters. *Ann. Statist. 7*, 756-771.

[16] Seidenfeld, T. (1981). On after-trial properties of best Neyman-Pearson confidence intervals. *Philosophy of Science 48*, 281-291.

[17] Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist. 26*, 243-258.

[18] Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist. 42*, 385-388.